REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
	Technical Report	-

4. TITLE AND SUBTITLE
IU Progress Report November 2013
Detecting Early Signatures of Persuasion in Information Cascades

5a. CONTRACT NUMBER: W911NF-12-1-0037
5b. GRANT NUMBER
5c. PROGRAM ELEMENT NUMBER
5d. PROJECT NUMBER
5e. TASK NUMBER
5f. WORK UNIT NUMBER

6. AUTHORS
A. Flammini, F. Menczer, Q. Mei, S. Malinchick

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Indiana University at Bloomington
Trustees of Indiana University
509 E 3RD ST
Bloomington, IN 47401 -3654

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
61766-NS-DRP.34

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
Indiana Universit team worked on two different core functions to refine the persuasion detection system that we are building: (i) the feature selection system for organic and promoted content classification; (ii) user influence detection.

University of Michigan team worked on i) a formal definition of rumor and ii) (with the help of two human

15. SUBJECT TERMS
promoted content classification; user influence detection; early detection of rumors

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
UU	UU	UU	UU		Alessandro Flammini
					19b. TELEPHONE NUMBER: 812-856-1830

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

Report Title

IU Progress Report November 2013

ABSTRACT

Indiana Universit team worked on two different core functions to refine the persuasion detection system that we are building: (i) the feature selection system for organic and promoted content classification; (ii) user influence detection.

University of Michigan team worked on i) a formal definition of rumor and ii) (with the help of two human annotators to label some rumors from our Boston Marathon Explosion dataset and refine the codebook of rumor. The inter rater reliability is at first 0.46 and then improved to 0.6 after several rounds of modification of codebook.
We have also been working on improving the performance of our rumor detection system. With human annotators, we had some very preliminar evaluation of our system.

DARPA SMISC Project:
DESPIC: Detecting Early Signatures of Persuasion in Information Cascades

Teams:
Indiana University: A. Flammini (PI) and F. Menczer
University of Michigan: Qiaozhu Mei
Lockheed Martin Advanced Technology Laboratories (ATL): S. Malinchik

Progress Report – November 2013

On October 30, 2013 the three teams (IU, Michigan and ATL) held a half-day bi-annual teleconference to discuss the next steps in the project development, including integration of core functions developed by each team, and delivery of an integrated framework during the Q3/Q4 of the last year of the project.

IU: During November 2013 the Indiana University team worked on two different core functions to refine the persuasion detection system that we are building: (i) the feature selection system for organic and promoted content classification; (ii) user influence detection.

Regarding point (i), our classification system adopts different off-the-shelf classifiers, including decisions trees, ensemble methods and Support Vector Machines. We also implemented our ad hoc SAX-VSM classifier that proved to be the one with the better classification performance and the higher computational efficiency. The system builds on the generation of five classes of features, namely network structure and diffusion patterns (retweet, mention and hashtag co-occurrence networks), language, content and natural language features (including Part-of-Speech tagging), timing features (e.g., inter- and intra-event distributions), sentiment features (e.g., emotion scores, valence-dominance-arousal scores, polarity scores, mood and attitude scores, etc.) and finally user meta-data inferred features (e.g., geo-information, follower-followee stastics, etc.). A greedy algorithm determines the discriminant power of each feature during the classification by means of a K-fold cross-validation step. During this process, the contribution of each feature to the classification is evaluated independently. At the end of the cross validation the algorithm ranks all features according to their performance and selects the top N for classification purpose (N can be either fixed or determined automatically by imposing a threshold on the increment in classification accuracy obtained by adding further features). In Figure 1 we report an example of feature selection where content features convey the higher predictive power. Results are obtained with a window length of 32 datapoints (16h40m) and window offset of 32 datapoints (16h40m) before the trending point.

According to point (ii), we are now developing a system to easily estimate the role and influence of Twitter users in a given discussion topic. The algorithm we devised is able to systematically categorize individuals in four different classes: (a) common users; (b) breadcasters; (c) influentials; and, (d) hidden influential users. To do so, the algorithm explores two dimensions: the ratio F of followers-followees that each user has, and the ratio M of mentions received by, versus those provided to other users. A user with low values of F and M will be considered a common user since it has generally less followees than followers and he/she provides more mentions than those he/she receives. The opposite conditions apply to identify influential users. The two non obvious classes are represented by broadcasters (i.e., users with more followers than followees) and hidden influential users (i.e., those who are disproportionately more mentioned with respect to what expected by their limited follower-followee ratio). We applied our system to characterize the users involved in the discussion around the Turkish protest know as *Gezi Park* to discover hidden influential and broadcaster users (see Figure 2). We plan to extend our approach to identify other potentially interesting users' behaviors this way, even in a dynamic context of networks and/or topics evolving over time.

Figure 1: Top ten most discriminant features between an organic trend (left) and a promoted one (right).

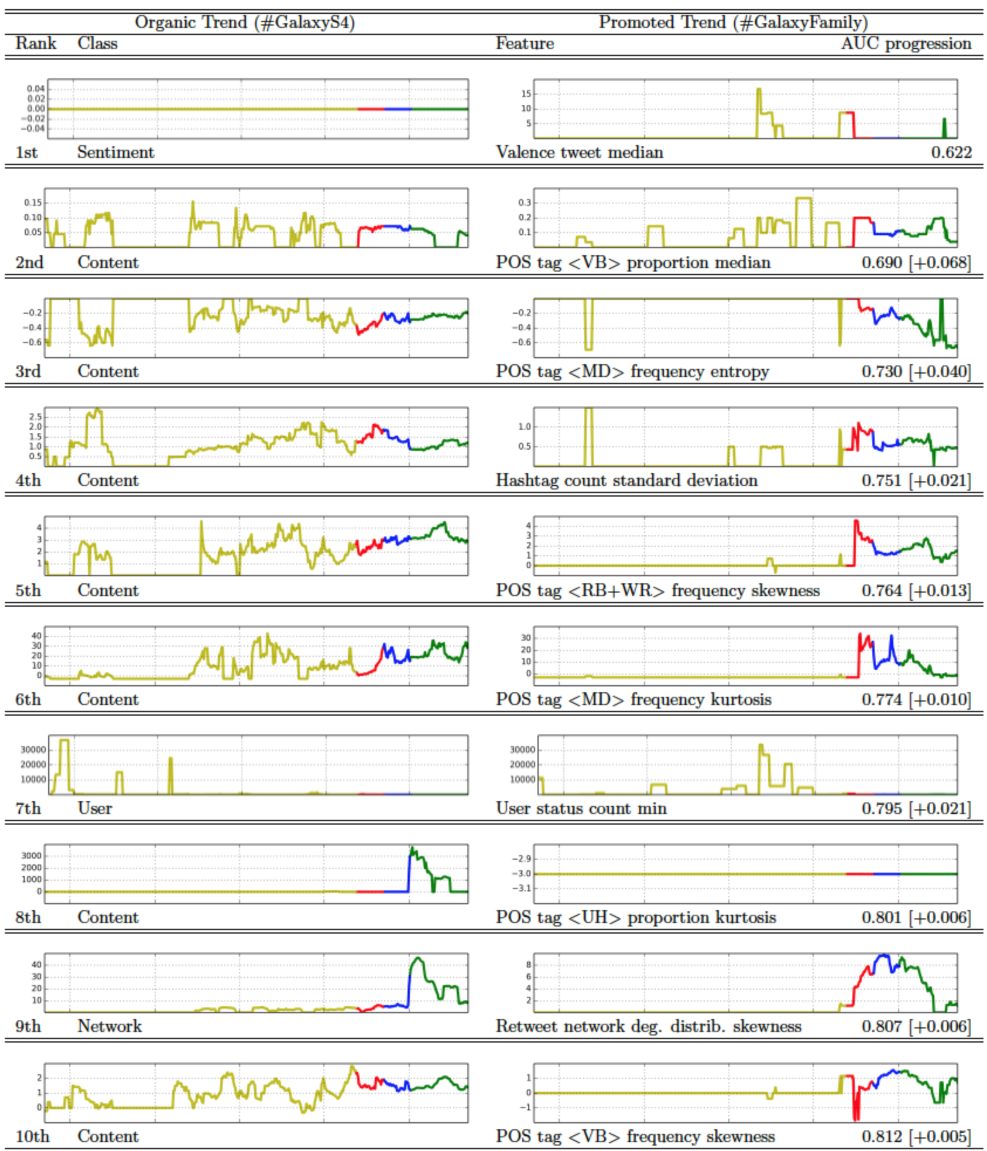
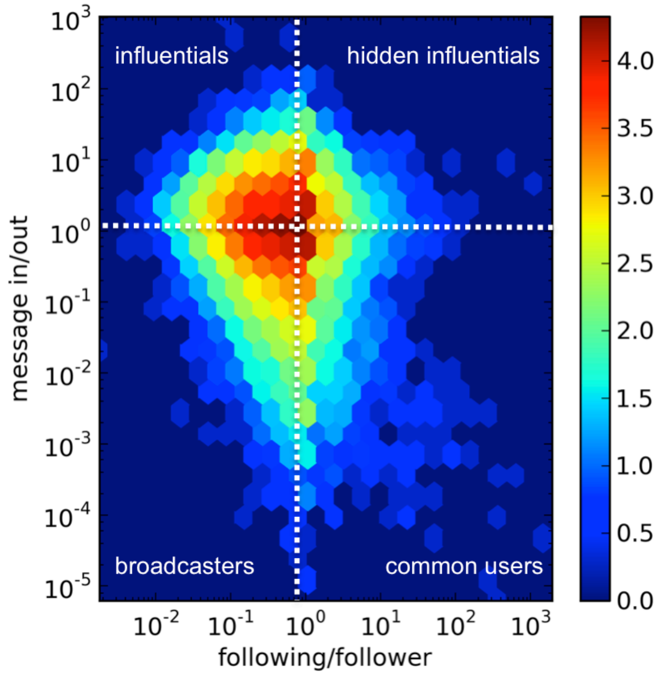


Figure 2: Map of user roles in the discussion about a given topic, as a function of their followers/followee ratio and in/out mentions.



UM: During the past month, we have been working on the early detection of rumors in Twitter. We define rumors in social media as statements that are wide spreading, controversial and fact-checkable. Based on this understanding about rumors, we found that usually at the early stage of the spread of a rumor, people will ask questions to express their suspicions, surprise, or effort to fact-check the rumor. Therefore, we construct a rumor detection system that consists of three parts: a question & correction filter that filters raw tweets that are in forms of questions and corrections, a statement detector that group tweets that are articulating the same statement together as one unique statement, and a statement assessment component that ranks the statements detected based on how likely they are rumors.

In the past month, we first had a careful discussion on how to formally define a rumor. We have been working with two human annotators to label some rumors from our Boston Marathon Explosion dataset and refine the codebook of rumor. The inter rater reliability is at first 0.46 and then improved to 0.6 after several rounds of modification of codebook. We have also been working on improving the performance of our rumor detection system. With human annotators, we had some very initial evaluation of our system.

In particular, we improved our rumor detection system. We refined the model and adopted different minhash and clustering algorithms to better detect statements after filtering the questions and corrections. We developed a summarization algorithm to summarize the statement from a cluster of tweets, then we used this statement to retrieve actual tweets that are not only questions and corrections in our data set. So that the number of tweets we retrieved will be used to rank the statement as potential rumors as output. After asking human annotators to label our system output, we reached a precision around 0.7 of top rumors detected from the 30 million tweets Boston Marathon Explosion dataset.

To handle large scale tweet stream, such as our tweets extracted from gardenhose API, we need to make our system capable of dealing with stream data. In order to achieve this, we rewrote the minhash and clustering algorithm, so that they can be run on mapreduce. We are also making plans on developing stream clustering algorithm for our system.

In the next month, we will continue training human annotators to help us modify our codebook. We are aiming at improving the inter rater reliability to 0.7. We will test our new version of minhash and clustering method, and then design a stream clustering algorithm to improve our system to handle stream data.

ATL: During the past month, ATL team has been working on validation of the code of our novel SAX-VSM classification method and testing different options of optimal parameter search strategy. We created also a presentation of our paper: Senin, P., Malinchik, S., "SAX-VSM: Interpretable Time Series Classification Using SAX and Vector Space Model" for 2013 IEEE 13th International Conference on Data Mining (ICDM'13) to be held on 7-10 December 2013 in Dallas, Texas, USA.